Steve Depp
413 – 55

Assignment 8
19 August 2019

1.1 <u>Use EDA to justify a log and or 1st difference transformation.</u>

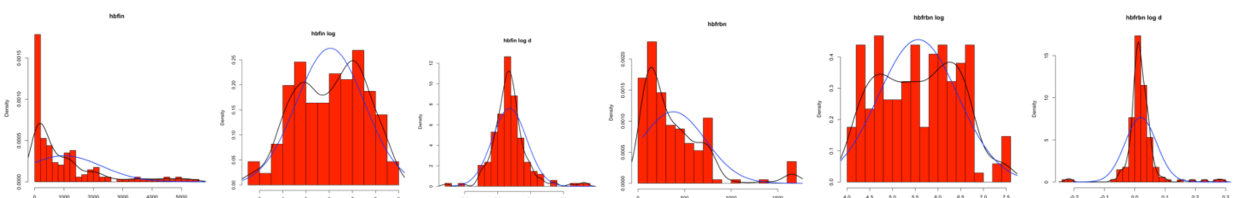
The following presents sample size, mean, standard deviation, median, 20% trimmed mean, min, max, range, skewness, excess kurtosis for time series of Fed debt held by foreign investors (hbfin) in the 1st three rows and Fed Reserve banks (hbfrbn) in the last 3 rows over 171 trading days taken every 3 months from Jan-1970 to July-2012. EDA first confirms there are no gaps in the 171 observations and converts dates from YYYY and MM format to monthly percent of year format for plotting. Data are log transformed in row 2 and 5 and log transformed and 1st differenced in rows 3 and 6.

	n	mean	sd	median	trimmed	min	max	range	skew	kurtosis
hbfin	171	1.005758e+03	1.264636e+03	502.00000000	728.02700730	12.4000000	5455.0000000	5442.6000000	1.82292077	2.7713973
hbfin log	171	6.050609e+00	1.461089e+00	6.21860012	6.07333025	2.5176965	8.6042879	6.0865914	-0.19717407	-0.8897172
hbfin log d	170	3.580348e-02	5.239707e-02	0.03170498	0.03301810	-0.1564128	0.2662271	0.4226399	0.95404417	4.7244526
hbfrbn	171	3.785731e+02	3.461179e+02	247.50000000	323.00437956	55.8000000	1664.7000000	1608.9000000	1.89392604	4.2138027
hbfrbn log	171	5.566256e+00	8.758905e-01	5.51141058	5.55815666	4.0217739	7.4174002	3.3956263	0.08824458	-0.9990276
hbfrbn log d	170	1.990532e-02	5.204355e-02	0.01562672	0.01744318	-0.2253063	0.2878344	0.5131407	0.98481779	12.3207297

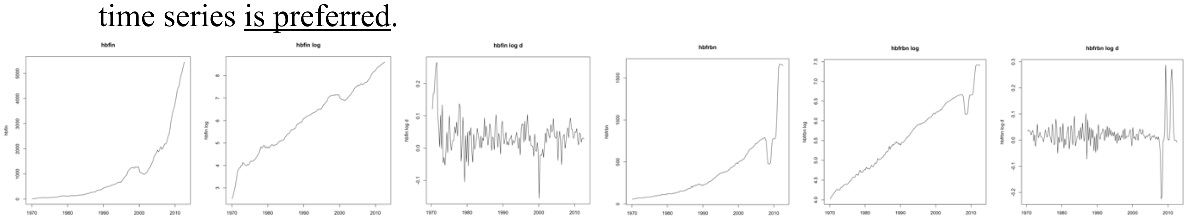
Untransformed data in rows 1 and 4 **above** have max that are orders of magnitude larger than min. A stable mean will be difficult to find in these series. Log transformed data have reduced skew and kurtosis versus both untransformed and log and 1st difference transformed series data. Of the 3 data forms, log and 1st differenced data have means closest to zero, greatest kurtosis, smallest variance and skew midway between that of the untransformed and log transformed.

It is worth noting that a t-test of the sample means of all series returned p-values that allow rejection of the null hypothesis of zero mean at the highest of confidence levels. For the 1st differenced log transformed series this is not surprising considering the standard deviation of the series is reduced by so much that the standard error of this t-test would be very small indeed.

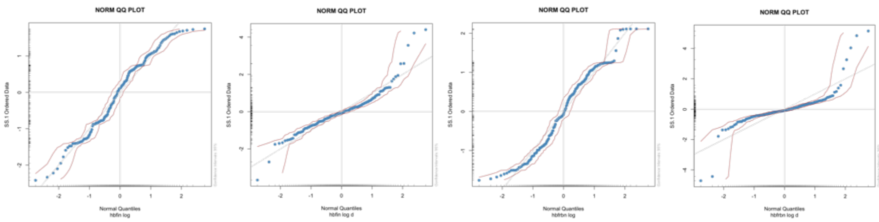
Thus univariates help us exclude the untransformed series, but not choose between the 2 alternative forms of transformation.



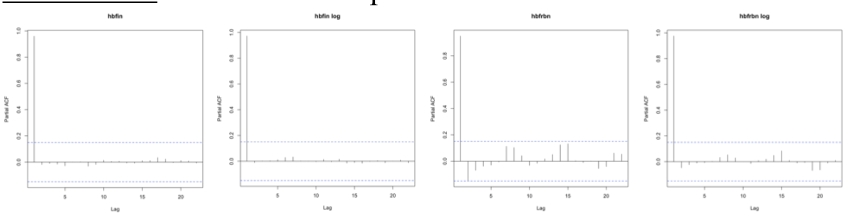
Histograms ordered same as the table **above** show log normal distributions (right skewed) in the untransformed series, reduced skew but multimodal log transformed series and highly kurtotic <u>1st differenced log transformed</u> data. From these plots, the latter form of time series <u>is preferred</u>.



Time series plots **above** ordered the same as above demonstrate a trending mean for untransformed and log transformed series and a stable mean and more apparent but heteroscedastic variance in the 1st differenced log transformed time series. A stable mean is preferred for stationarity; heteroscedasticity is not, but the 1st differenced series are preferred.

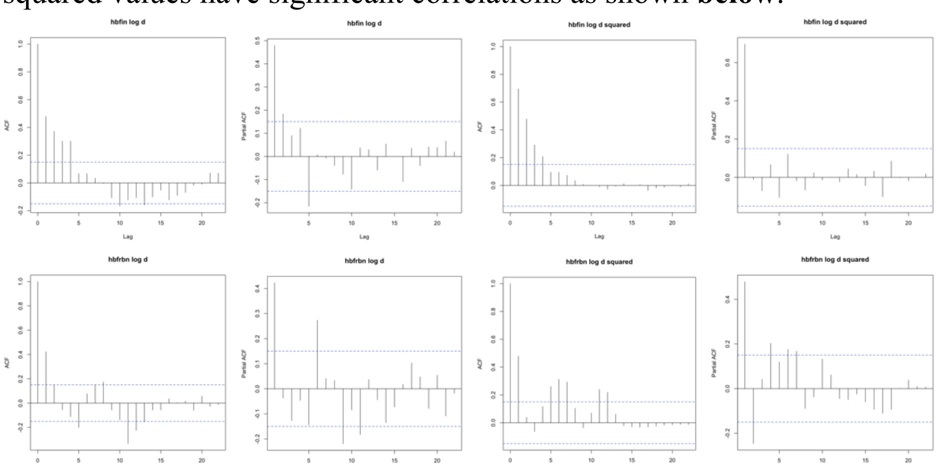


QQ plots **above** leave out the untransformed series (but ordered the same as above) represent the enhanced skew, fatter tails and outliers generated when log transformed series (plots 1 and 3 above) are 1st differenced. One might prefer the undifferenced log transformed data from these plots.



Augmented Dickey Fuller tests p-values = 0.99 for all 6 series (not shown) do not allow rejection of null hypothesis of unit roots in these data. KPSS test statistics for all 6 series (not shown) are consistent with ADF tests allowing rejection of null hypothesis of no unit roots. Repeated KPSS test of all 6 series and their squared values (12 series in total) show 1st differencing needed in all cases except the three 1st differenced log transformed unsquared series. So, KPSS tests of the 1st differenced log transformed squared series indicates GARCH effects. The ‘hbfin’ series shows 2nd differencing may be needed for both unsquared and squared versions of untransformed series. Partial ACF plots **above** for the untransformed and log transformed series show a significant partial auto correlation at the 1st lag. Unit roots would suggest 1st differenced series is preferred.

Portmanteau test of autocorrelation via Box-Ljung has p-value allowing rejection of the hypothesis of no serial correlations at the highest of confidence levels for all series. For the 1st difference of the log transformed series, the ACF PAC plots for unsquared and squared values have significant correlations as shown **below**.



Plots **above** are for 1st differenced log tranformed series. On **top** are ACF and PACF for unsquared on **left**, ACF and PACF for squared values of "hbfin" on **right**; on **bottom** are the same for "hbfrbn". None suggest independence. The exponential decline in ACF and PACF of unsquared "hbfin" suggest ARMA process. The exponential decline in ACF and sharp drop off in PACF at lag 1 for squared "hbfin" suggests an AR process or ARCH effects at lag 1. The unsquared and squared "hbfrbn" series ACF and PACF show more complexity: there is initial significant correlation decline followed by an opposite sign significant correlation for all but ACF of squared values. More investigation is needed.

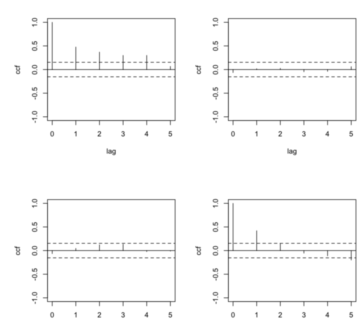
	n	mean	sd	median	trimmed	min	max	range	skew	kurtosis
hbfin sd	10	176.44538057	276.81275780	76.91875212	103.27595848	7.69744037	930.5486975	922.85125713	1.943090	2.551831
hbfin log sd	10	0.20901737	0.13474139	0.19059072	0.18478653	0.05990753	0.5519739	0.49206641	1.456597	1.386492
hbfin log d sd	10	0.04341679	0.02434943	0.03474239	0.03854245	0.02145887	0.1043695	0.08291061	1.419973	1.063770
hbfrbn sd	10	74.05665243	146.43130574	30.07601079	30.84112812	7.03451574	486.8029836	479.76846781	2.196669	3.330515
hbfrbn log sd	10	0.13592912	0.12781646	0.10521257	0.10208102	0.05184921	0.4907939	0.43894465	2.072304	2.980029
hbfrbn log d sd	10	0.04159157	0.03381043	0.03059628	0.03434964	0.01728890	0.1238297	0.10654076	1.392781	0.685140

The **above** provides univariates after dividing the series into 10 chronological segments and taking each segments standard deviation. For both series, the standard deviation of standard deviation is least for the 1st differenced log transformed series, declining from 77 for "hbfin" and 30 for "hbfrbn" to 0.03 for both in their 1st differenced log transformed state. One can see from the range that the variance of variance is substantially reduced via log transformation and then marginally reduced via 1st differencing.

```
        Pearson's product-moment correlation

data:  debt_logd$hbfin and debt_logd$hbfrbn
t = -0.83397, df = 168, p-value = 0.4055
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.21266810  0.08714732
sample estimates:
        cor
-0.06420915
```

```
[1] "Covariance matrix:"
          hbfin    hbfrbn
hbfin   0.002745 -0.000175
hbfrbn -0.000175  0.002709
CCM at lag:  0
        [,1]    [,2]
[1,]  1.0000 -0.0642
[2,] -0.0642  1.0000
Simplified matrix:
CCM at lag:  1
+ .
. +
CCM at lag:  2
+ .
. .
CCM at lag:  3
+ .
. .
CCM at lag:  4
+ .
. .
CCM at lag:  5
. .
. -
```

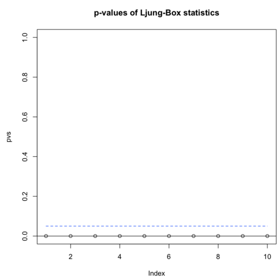


1.2 With a p-value of 0.4055, the output on the **left above** represents that concurrent correlation between the two 1st differenced log transformed time series is zero since one cannot reject the null hopothesis of zero correlation. This makes sense since the correlation at 6.4% is slight.

In the **center above** output are the cross correlation matrices (CCM) for lag 0 through 5. The simplified cross correlations with dots in the off diagonal cells for lags 1 to 5 indicate that the 2 series are decoupled. In the **above right** output, that same decoupling is indicated by the independence shown in the **upper right** and **lower left** CCF plots at lags 1 to 5. Lag 0 in the upper right and lower left echo the independence of concurrent correlation shown in the test of correlation discussed in the previous paragraph.

The (+) in upper left of CCM 1 – 4 indicate significant “hbfin” autocorrelation at lag 1 to 4 also exhbited in the upper left CCF with bars piercing the confidence band for lags 1 to 4. The (+) and (-) at lag 1 and 4 respectively in the lower right CCM are also seen in the lower right CCF with lag 1 and 4 lines through the confidence band on top and bottom respectively. These indicate significant “hbfrbn” positive and negative autocorrelation at lags 1 and 4. Finally, lag zero in the upper left and lower right CCF say that each series is 100% concurrently correlated with itself; this is also seen with 1.000 in the CCM at lag.

```
       m,  Q(m)    and p-value:
[1] 1.0000e+00 7.2450e+01 6.8834e-15
[1]   2.00 104.77   0.00
[1]   3.00 125.15   0.00
[1]   4.00 143.39   0.00
[1]   5.00 151.89   0.00
[1]   6.00 154.57   0.00
[1]   7.00 159.79   0.00
[1]   8.00 165.59   0.00
[1]   9.00 168.58   0.00
[1]  10.00 178.13   0.00
```



1.3 Shown by the Box-Ljung tests $Q_2(10)$ test statistic = 178.13 and p-value = 0, the null hypothesis H_0: $\rho_1 = \cdot\cdot\cdot = \rho_{10} = 0$ is rejected at the highest confidence level. Alternate hypothesis $H_{a:}$ ρ_i not = zero. Tsay on page 397 says this “statistic is used to test that there are no auto- and cross correlations in the vector series”. We have 2 series that are uncoupled but have significant autocorrelation. This result is consistent with output found in problem 1.2.

2.1 **Use EDA to justify a VAR(4) model.**

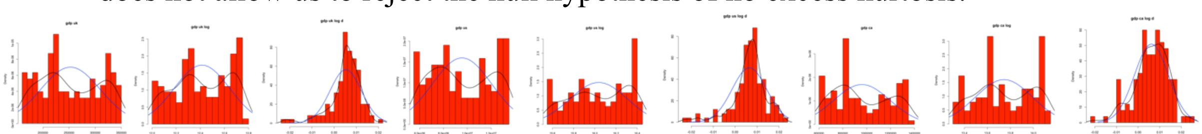
Please bear with me as I’ve found that EDA for log transformation and 1st differences also helps to identify other characteristics of the series that help justify model order.

The following presents sample size, mean, standard deviation, median, 20% trimmed mean, min, max, range, skewness, excess kurtosis for time series of GDP data for the UK in the 1st set of three rows, the US in the 2nd set of three rows and Canada in the final set of three rows over 126 quarterly reports from Jan-1980 to Apr-2011. EDA first confirms there are no gaps in the 126 observations and converts dates from YYYY and MM format to monthly percent of year format for plotting. Data are log transformed in row 2, 5 and 8 and log transformed and 1st differenced in rows 3, 6 and 9.

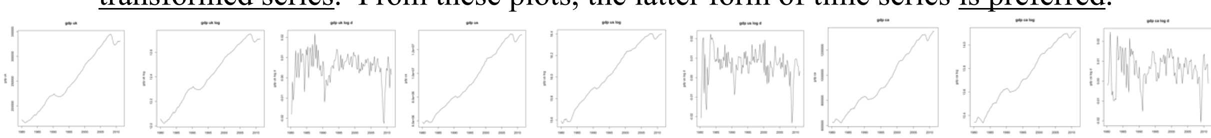
	n	mean	sd	median	trimmed	min	max	range	skew	kurtosis
gdp uk	126	2.525709e+05	5.689382e+04	2.431690e+05	2.523008e+05	1.660520e+05	3.448090e+05	1.787570e+05	0.11362748	-1.362611
gdp uk log	126	1.241362e+01	2.298739e-01	1.240151e+01	1.242028e+01	1.202006e+01	1.275075e+01	7.306896e-01	-0.12058156	-1.286579
gdp uk log d	125	5.223092e-03	7.086442e-03	6.102259e-03	5.953169e-03	-2.250303e-02	2.173321e-02	4.423624e-02	-1.31932426	2.966034
gdp us	126	9.551488e+06	2.497971e+06	9.152500e+06	9.551779e+06	5.776600e+06	1.332600e+07	7.549400e+06	0.07497171	-1.411987
gdp us log	126	1.603674e+01	2.706807e-01	1.602953e+01	1.604710e+01	1.556933e+01	1.640523e+01	8.359017e-01	-0.18479995	-1.301279
gdp us log d	125	6.473996e-03	7.872912e-03	7.453007e-03	7.092607e-03	-2.327649e-02	2.222757e-02	4.550406e-02	-1.12347839	2.512260
gdp ca	126	9.598270e+05	2.356097e+05	8.998880e+05	9.552897e+05	6.201160e+05	1.350078e+06	7.299620e+05	0.22542414	-1.359425
gdp ca log	126	1.374425e+01	2.479603e-01	1.371002e+01	1.374830e+01	1.333766e+01	1.411567e+01	7.780111e-01	-0.01635098	-1.323562
gdp ca log d	125	6.153672e-03	7.851955e-03	6.839232e-03	6.654755e-03	-2.067693e-02	2.467191e-02	4.534884e-02	-0.58290554	0.548218

Untransformed data in rows 1 and 4 **above** have the greatest standard deviation of the 3 data states. Log transformation reduces the mean and standard deviation and shifts distribution skew from right to left around symmetric. 1^{st} differencing further reduces mean and standard deviation and increases skew toward the left (negative) and increases excess kurtosis. Of the 3 data forms, the data transformed by logarithm and 1^{st} differencing have means closest to zero, greatest kurtosis, smallest variance, but largest absolute value skew.

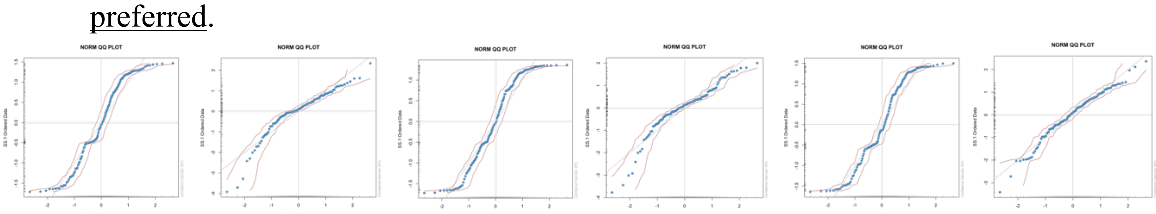
Despite the above univariates, test appears to indicate the untransformed and log transformed series are symmetric and 1^{st} differenced log transformed series means are not zero which does not allow for preference based on those moments. Results of tests not shown: Thus, for all forms of all series, a t-test of the sample means returned p-values that allow rejection of the null hypothesis of zero mean at the highest of confidence levels. For the untransformed and log transformed forms of series, a sample test cannot reject null hypothesis of symmetry. A test of 1^{st} differenced log form of series rejects the null hypothesis of symmetry at the highest level for UK and US series and at the 99.9% level for the Canada series. A test of all series and forms except the Canada 1^{st} differenced log transformed allows rejection of the null hypothesis of no excess kurtosis at the highest confidence level. The 1^{st} differenced log transformed Canada series test does not allow us to reject the null hypothesis of no excess kurtosis.



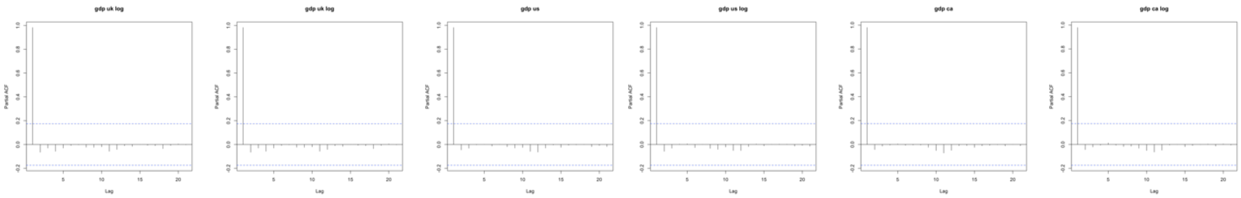
Histograms ordered same as the table **above** show distributions that are not normal (are uniform-ish?) for untransformed and log transformed series and distinctly more bell shaped and kurtotic distributions with left side outliers for the 1^{st} differenced log transformed series. From these plots, the latter form of time series is preferred.



Although tiny in size, the plots of time series **above** ordered the same as the above table exhibit similarities between untransformed and log transformed, which have non-constant means and between countries such that the broader patterns of GDP seem similar. There is heteroscedasticity apparent in the 1^{st} differenced log transformed time series which is also at similar points in time between countries. With sharp and extended drops in these 1^{st} differenced series, it might be that stable means are questionable, but means are surely more stable with these series than untransformed and log transformed. A stable mean is preferred for stationarity; heteroscedasticity is not, but the 1^{st} differenced series are preferred.

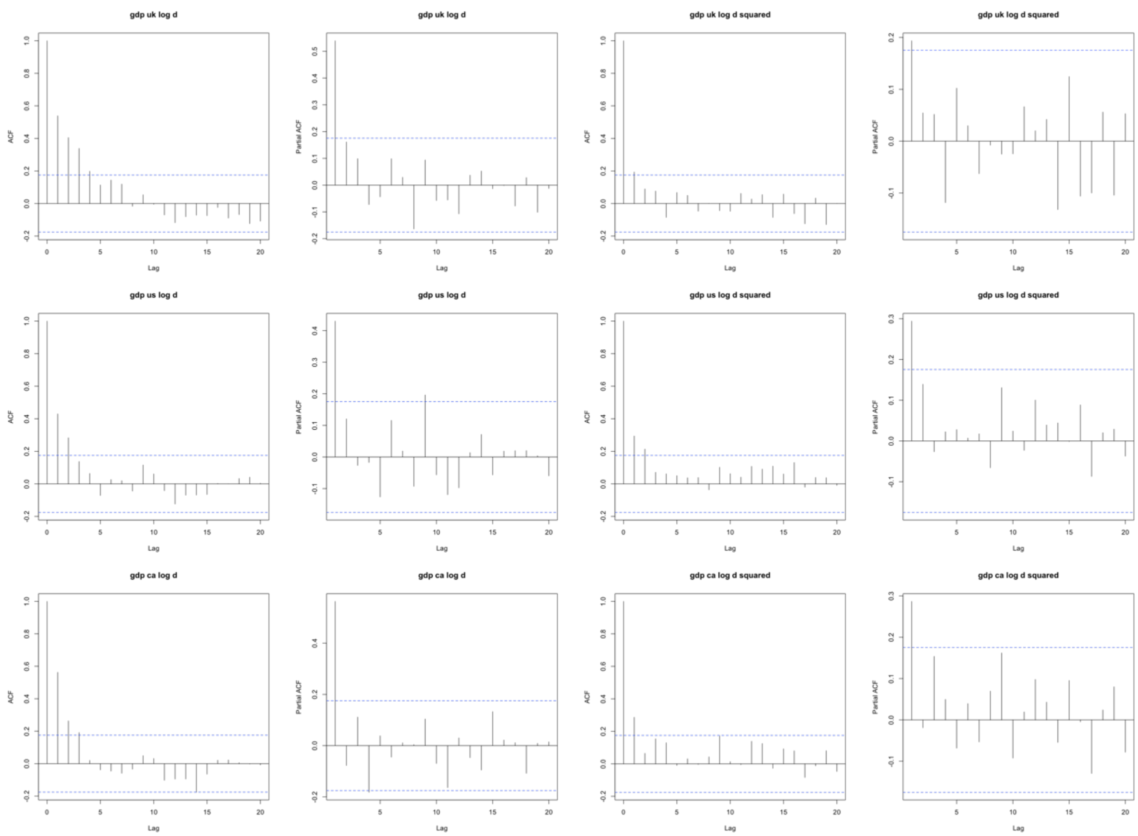


QQ plots of 1^{st} differenced log transformed series **above positions 2, 4 and 6** represent left skewed and kurtotic distributions with outliers, and seem to improve upon the log transformed series that are cut off at 1.5 SDs; those points normally beyong 1.5 SDs appear to collect in the shoulders. One might <u>prefer the undifferenced log transformed</u> data from these plots.



Augmented Dickey Fuller tests p-values for the untransformed and log transformed series do not allow rejection of the null hypothesis of unit roots. Same tests for 1st differenced log transformed series allow rejection of null hypothesis at at least 90% confidence level, but some at the 99.9% confidence level, depending upon whether the test is for series with constant or not. KPSS test statistics provide similar results with test statistics that allow rejection of the null hypothesis of no unit roots at the 99.9% level and repeated KPSS tests suggesting existence of 1 unit root in the unsquared and 2 in the squared data. After 1st differencing, the ADF and KPSS suggest no unit roots in all 3 unsquared series, and the UK squared series. KPSS tests of the squared US and Canada series does not allow rejection of the null hypothesis for unit roots and suggests a 1st differencing. Thus, for these 2 series, GARCH effects may be present. PACF plots **above** for the six untransformed and log transformed series represent significant partial autocorrelation at 1 period lag. All taken, 1st differencing appears to be warranted for the unsquared series.

Portmanteau test of autocorrelation via Box-Ljung has p-value allowing rejection of hypothesis of no serial correlation at the highest of confidence levels for all 9 of the unsquared series (3 countries x 3 forms) and for the untransformed and log transformed squared series. The squares of 1st differenced log transformed data does not allow rejection of the null hypothesis for no serial correlation for UK series, but allows rejection in the case of US and Canadian series at the 95% confidence level. This last result is consistent with the KPSS results in the previous paragraph. At the very least, the squares of 1st differenced data require investigation.



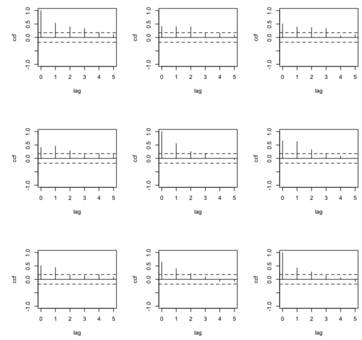
Plots **above** are for 1st differenced log tranformed series. Top row UK, middle US, bottom Canada. On **left** are ACF and PACF for unsquared; on **left** are ACF and PACF for squared series. Generally, all ACF are exponentially declining and all PACF stop at 1 lag suggesting an AR(1) and ARCH(1) model. One can discern significant PACF at lag 4 for Canada and lag 9 for US, both unsquared, but these could easily be due to chance.

	n	mean	sd	median	trimmed	min	max	range	skew	kurtosis
gdp uk sd	10	6.401934e+03	2.869567e+03	7.316094e+03	6.643842e+03	1.696255e+03	9.172342e+03	7.476087e+03	-0.52154251	-1.5770709
gdp uk log sd	10	2.586510e-02	1.228283e-02	2.804037e-02	2.587462e-02	7.669477e-03	4.398455e-02	3.631508e-02	-0.15138879	-1.5810296
gdp uk log d sd	10	5.273951e-03	2.709009e-03	5.053909e-03	5.019536e-03	2.695585e-03	9.887638e-03	7.192054e-03	0.56538271	-1.2840079
gdp us sd	10	2.748137e+05	1.138571e+05	2.764375e+05	2.770641e+05	7.556372e+04	4.560608e+05	3.804971e+05	-0.10500389	-1.0688444
gdp us log sd	10	2.960157e-02	1.384166e-02	2.826721e-02	2.859910e-02	1.280590e-02	5.441698e-02	4.161108e-02	0.40224716	-1.3757912
gdp us log d sd	10	6.139583e-03	3.142289e-03	5.300023e-03	5.558916e-03	3.349727e-03	1.357477e-02	1.022505e-02	1.25194152	0.3671649
gdp ca sd	10	2.818693e+04	1.142203e+04	3.292844e+04	2.874502e+04	8.682374e+03	4.322676e+04	3.454438e+04	-0.58450372	-1.2177271
gdp ca log sd	10	3.015917e-02	1.380088e-02	3.077306e-02	3.008297e-02	1.061264e-02	5.031523e-02	3.970258e-02	0.02023983	-1.6708018
gdp ca log d sd	10	6.441578e-03	2.469718e-03	5.956054e-03	6.150077e-03	3.584461e-03	1.163070e-02	8.046241e-03	0.68554649	-0.6854997

The **above** provides univariates after dividing the series into 10 chronological segments and taking each segments standard deviation. For all country series, the standard deviation of standard deviation is least for the 1st differenced log transformed series.

```
[1] "Covariance matrix:"
         uk       ca       us
uk 5.02e-05 2.28e-05 2.83e-05
ca 2.28e-05 6.17e-05 3.99e-05
us 2.83e-05 3.99e-05 6.20e-05
CCM at lag:  0
      [,1]  [,2]  [,3]
[1,] 1.000 0.409 0.507
[2,] 0.409 1.000 0.645
[3,] 0.507 0.645 1.000
```

```
Simplified matrix:
CCM at lag:  1
+ + +
+ + +
+ + +
CCM at lag:  2
+ + +
+ + +
. + +
CCM at lag:  3
+ . +
. . .
. . .
CCM at lag:  4
. . .
. . .
. . .
CCM at lag:  5
. . .
. . .
. . .
```



To justify the use of a VAR(4) model we would use cross correlation matrices CCM and cross correlation function plots on the 1st differenced log transformed data. As with problem 1.2 and 1.3, we view simplified CCM and CCF plots **above** to confirm serial dependence in the multivariate series at the 5% significance level**.** Specifically, significant autocorrelations are seen in the CCM diagonal and CCF diagonal plots for all 3 country time series out to lag 2 and for UK vs US out to lag 3. There is significant feedback amongst all 3 country series at lag 1 and amongst all country series but the US at lag 2. At lag 2, US has a unidirectional relationship with the UK where the US series at lag 2 leads UK series. The US series has a feedback relationship with the Canada series in lag 2. In lag 3 CCM and in the upper right CCF plot, we see the UK series is seen led by lag 3 US series: this relationship is unidirectional.

```
        m,  Q(m)    and p-value:
[1] 1.0000e+00 7.9140e+01 2.3948e-13
[1] 2.0000e+00 1.1810e+02 1.1102e-16
[1]   3.00 143.31   0.00
[1]   4.00 163.94   0.00
[1]   5.00 174.66   0.00
[1] 6.0000e+00 1.8500e+02 3.3307e-16
[1] 7.0000e+00 1.9982e+02 3.3307e-16
[1] 8.0000e+00 2.1409e+02 5.5511e-16
[1] 9.0000e+00 2.1973e+02 9.5479e-15
[1] 1.0000e+01 2.2669e+02 8.9928e-14
```

There does not appear to be evidence for modeling a VAR(4) series from the CCM / CCF plot analysis. Multivariate Box-Ljung portmanteau test result p-values = 0 **above** represent that we can reject the null hypothesis of independence in the CCMs tested to lag 10. Since we were unable to obtain significant autocorrelations or cross correlations past lag 3, the contradiction of these two tests might suggest cointegration.

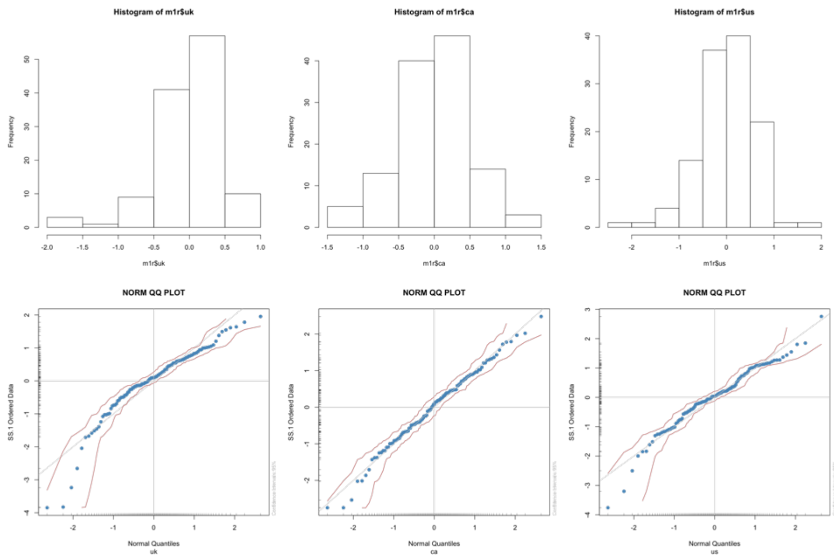
2.2 Fit a VAR(4) model to the series and perform model checking. Where z_t is composed of the 1st differenced log transformed times series of UK CA and US GDP data.

$$z_t = \begin{bmatrix} 0.148 \\ 0.0776 \\ 0.239 \end{bmatrix} + \begin{bmatrix} 0.516 & 0.0719 & 0.0639 \\ 0.378 & 0.3160 & 0.4096 \\ 0.519 & 0.1730 & 0.1504 \end{bmatrix} z_{t-1} + \begin{bmatrix} -0.0504 & 0.160 & -0.00198 \\ -0.174 & -0.254 & 0.06295 \\ -0.2178 & -0.159 & 0.22561 \end{bmatrix} z_{t-2} +$$
$$\begin{bmatrix} 0.0524 & -0.2788 & 0.1411 \\ 0.0962 & 0.1203 & 0.0137 \\ 0.0478 & -0.0786 & 0.0738 \end{bmatrix} z_{t-3} + \begin{bmatrix} 0.0401 & 0.2617 & -0.2465 \\ 0.0747 & -0.0903 & -0.0978 \\ 0.1541 & -0.1518 & -0.0535 \end{bmatrix} z_{t-4} + a_t$$

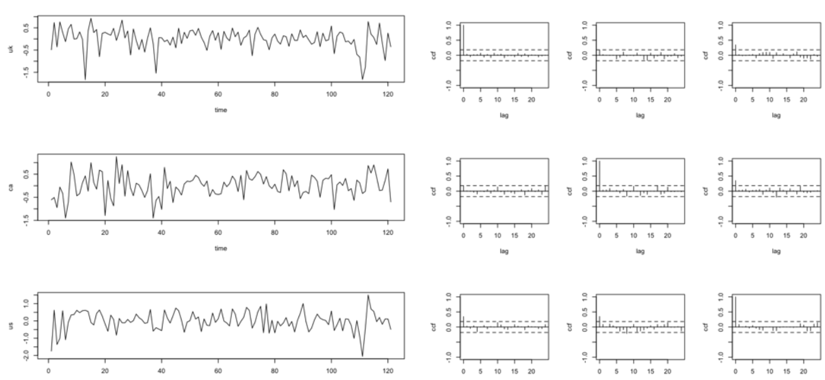
With a residual covariance matrix $\begin{bmatrix} 0.2243 & 0.0438 & 0.0893 \\ 0.0438 & 0.2569 & 0.0968 \\ 0.0893 & 0.0968 & 0.2955 \end{bmatrix}$

And AIC = -3.761, BIC = -2.947, HQ = -3.431

	n	mean	sd	median	trimmed	min	max	range	skew	kurtosis
uk	121	-9.762793e-17	0.4755768	0.04581094	0.04610992	-1.830130	0.928976	2.759106	-1.3058027	3.0098123
ca	121	-3.971388e-16	0.5089691	0.04060807	0.01203631	-1.398317	1.261642	2.659958	-0.2739936	0.1227212
us	121	4.156341e-16	0.5458428	0.01850695	0.02919164	-2.052769	1.502941	3.555710	-0.6582204	1.5124436



Above are univariates of UK, CA, US residuals appearing to represent varying degrees of negative skew and leptokurtosis that are similar in rank order (UK, CA, US) to those of the original data that were also negatively skewed and leptokurtotic. In fact, however, single sample tests of **<u>means,</u>** skew and kurtosis of all 3 distributions are unable to reject the null hypothesis of **<u>zero</u>** values for these descriptors. Histograms **above** show left skew for UK and US and symmetry for CA, QQ plots suggest one source of negative skew might be a set of 5-6 outliers for each series.



Ljung-Box Statistics:

	m	Q(m)	df	p-value
[1,]	1.00	1.50	9.00	1.00
[2,]	2.00	2.32	18.00	1.00
[3,]	3.00	5.58	27.00	1.00
[4,]	4.00	10.75	36.00	1.00
[5,]	5.00	15.29	45.00	1.00
[6,]	6.00	20.88	54.00	1.00
[7,]	7.00	27.09	63.00	1.00
[8,]	8.00	40.83	72.00	1.00
[9,]	9.00	43.21	81.00	1.00
[10,]	10.00	54.44	90.00	1.00
[11,]	11.00	66.00	99.00	1.00
[12,]	12.00	78.49	108.00	0.99
[13,]	13.00	85.51	117.00	0.99
[14,]	14.00	93.95	126.00	0.99
[15,]	15.00	99.88	135.00	0.99
[16,]	16.00	105.72	144.00	0.99
[17,]	17.00	113.25	153.00	0.99
[18,]	18.00	121.31	162.00	0.99
[19,]	19.00	133.74	171.00	0.98
[20,]	20.00	145.06	180.00	0.97
[21,]	21.00	148.91	189.00	0.99
[22,]	22.00	155.25	198.00	0.99
[23,]	23.00	161.43	207.00	0.99
[24,]	24.00	177.06	216.00	0.98

Time series plots of residuals **above left** suggest **<u>constant variance</u>** and **<u>constant mean</u>**, and CCF plots **above center** (and associated CCM not shown) indicate **<u>independence</u>** by way of no autocorrelations or cross correlations that are significant at the 0.05 level. Box-Ljung statistics in the table **above right** suggest **<u>independence</u>** as well. A Box-Ljung test of each time series residuals squared (**not shown**) (which my reading suggests is equivalent to a McLeod-Li test of residuals for constant variance) returns p-values suggesting we cannot reject the null hypothesis of **constant variance**.

Thus the model is adequate, although some parameters appear not to be significant when compared with their stanadard errors.

2.3 Simplify the model by removing insignificant parameters with type-1 error rates at α = 0.05. Where I believe $z_t = (\, z_{uk,t}\; z_{ca,t}\; z_{us,t}\,)'$.

$$z_t = \begin{bmatrix} 0.204 \\ 0 \\ 0.323 \end{bmatrix} + \begin{bmatrix} 0.565 & 0.000 & 0.000 \\ 0.376 & 0.299 & 0.408 \\ 0.508 & 0.268 & 0.000 \end{bmatrix} z_{t-1} + \begin{bmatrix} 0 & 0.284 & 0 \\ 0 & -0.153 & 0 \\ 0 & 0.000 & 0 \end{bmatrix} z_{t-2} +$$

$$. \quad \begin{bmatrix} 0 & -0.249 & 0 \\ 0 & 0.000 & 0 \\ 0 & 0.000 & 0 \end{bmatrix} z_{t-3} + \begin{bmatrix} 0 & 0.296 & -0.243 \\ 0 & 0.000 & 0.000 \\ 0 & -0.199 & 0.000 \end{bmatrix} z_{t-4} + a_t$$

With a residual covariance matrix $\begin{bmatrix} 0.237 & 0.048 & 0.101 \\ 0.048 & 0.278 & 0.11 \\ 0.101 & 0.11 & 0.329 \end{bmatrix}$

Where now due to zeroed out parameter components:

At lag 1: UK is not led CA or US.
US has no AR component with itself.
At lag 2: Only UK is led by CA and CA has it's autoregression component.
At lag 3: Only UK is led by CA.
At lag 4: UK is led by CA and US.
US is led by CA and retains its autoregressive component.

Which may be easier for some to see in this format: z_1 = UK, z_2 = CA z_3 =US

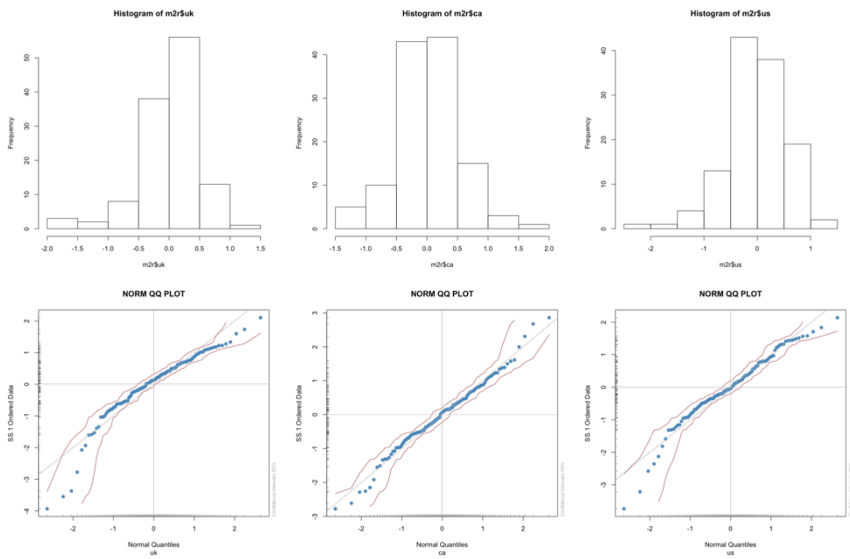
$z_{1t} = 0.204 + 0.565_{z1,t-1} + 0.284_{z2,t-2} - 0.249_{z2,t-3} + 0.296_{z2,t-4} - 0.243_{z3,t-4} + a_{1t}$
$z_{2t} = 0.376_{z1,t-1} + 0.299_{z2,t-1} + 0.408_{z3,t-1} - 0.153_{z2,t-2} + a_{2t}$
$z_{3t} = 0.323 + 0.508_{z1,t-1} + 0.268_{z2,t-1} - 0.199 z_{2,t-4} + a_{3t}$

And AIC = -3.93, BIC = -3.66, HQ = -3.82

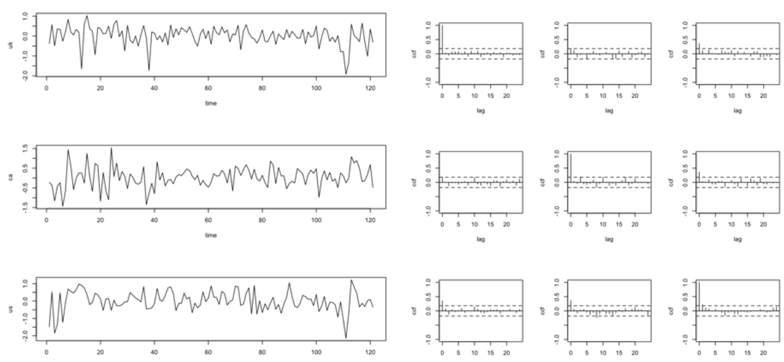
	n	mean	sd	median	trimmed	min	max	range	skew	kurtosis
uk	121	4.113360e-16	0.4886293	0.06282377	0.05000656	-1.920969	1.027784	2.948753	-1.31378134	2.9227852
ca	121	3.123975e-02	0.5281764	0.06853225	0.03791422	-1.435509	1.540085	2.975594	-0.06910388	0.6188036
us	121	7.872069e-17	0.5755883	-0.02413990	0.02896307	-2.150168	1.228882	3.379051	-0.72050973	1.4527640

uk H1: skew=zeroNULL	ca H1: skew=zeroNULL	us H1: skew=zeroNULL
-0.536348984762236	-0.028211539687721	-0.29414686753479
0.591717382068883	0.977493433548457	0.768645679008392
uk H1: kurtosis=zeroNULL	ca H1: kurtosis=zeroNULL	us H1: kurtosis=zeroNULL
0.59661102514149	0.126312750848384	0.29654421635326
0.550767108611122	0.899484363383972	0.76681450480553

In the residual univariates **shown above** we see zero means that via single sample t-tests do not allow us to reject the null hypothesis of zero means. Although skew and kurtosis appear large, tests of their significance do not allow us to reject the null hypothesis of symmetry and zero excess kurtosis. Variance levels in residuals are similar for each time series of residuals but notice the residual covariance matrix components are greater with the simplified model. This makes since less is determined by this model. The decline in parameter components (DOF) may be a factor in improved AIC BIC HQ measures as well.



One can see in the **above** plots the reduced skew in CA histogram and QQ plot as left side outliers are more along the normal line than in the VAR(4) model residuals and versus in the UK and US time series residuals of this VAR(2)



```
Hit Enter to compute MQ-statistics:
Ljung-Box Statistics:
         m       Q(m)     df    p-value
 [1,]   1.00     7.93    9.00     0.54
 [2,]   2.00    16.26   18.00     0.57
 [3,]   3.00    24.71   27.00     0.59
 [4,]   4.00    27.71   36.00     0.84
 [5,]   5.00    36.06   45.00     0.83
 [6,]   6.00    41.54   54.00     0.89
 [7,]   7.00    47.78   63.00     0.92
 [8,]   8.00    64.66   72.00     0.72
 [9,]   9.00    68.22   81.00     0.84
[10,]  10.00    81.99   90.00     0.71
[11,]  11.00    91.03   99.00     0.70
[12,]  12.00   101.83  108.00     0.65
[13,]  13.00   111.75  117.00     0.62
[14,]  14.00   119.75  126.00     0.64
[15,]  15.00   127.84  135.00     0.66
[16,]  16.00   134.71  144.00     0.70
[17,]  17.00   142.92  153.00     0.71
[18,]  18.00   150.03  162.00     0.74
[19,]  19.00   161.08  171.00     0.70
[20,]  20.00   172.29  180.00     0.65
[21,]  21.00   175.93  189.00     0.74
[22,]  22.00   180.17  198.00     0.81
[23,]  23.00   188.40  207.00     0.82
[24,]  24.00   204.69  216.00     0.70
```

```
        Box-Ljung test

data:  m1r$us^2
X-squared = 26.203, df = 10, p-value = 0.003476
```

Once again the residuals of the VAR model appear stationary and the model adequate. Time series plots of residuals **above left** suggest **<u>constant variance</u>** and **<u>constant mean</u>**, and CCF plots **above center** (and associated CCM not shown) indicate **<u>independence</u>** by way of no autocorrelations (except for US at lag 1) or cross correlations that are significant at the 0.05 level. Box-Ljung statistics in the table **above right** suggest **<u>independence</u>** as well. A Box-Ljung test of each time series residuals squared (which my reading suggests is equivalent to a McLeod-Li test of residuals for constant variance) returns p-values suggesting we cannot reject the null hypothesis of **constant variance for UK and CA but not US.** We can reject the null hypothesis of constant variance for US residuals sqared at the 99% confidence level (shown above).

Thus the model seems adequate.

3.0 Executive summary:

The model for predicting quarterly GDP reveals several interesting relationships.

We developed 2 models and the more accurate of the 2 as measured by several information ratios contains fewer parameter components. This has 2 benefits.

One is that the simpler model will generalize better when employed for forecasting in the future. Two is that the simpler model reveals which national GDP are more determined by other nations GDP. For example, the full model, whose predictions perform worse, includes references to the previous years' worth of GDP data for all nations.

The simple model for UK looks at one prior UK GDP, 3 prior Canada GDP, and 1 US GDP.

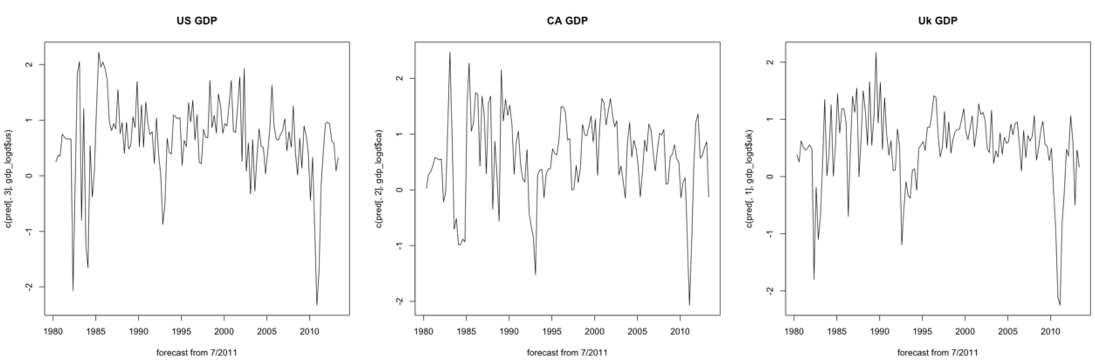
The simple model for CA looks at fewer components: two prior Canada GDP, and one each of UK and US GDP.

The US model looks at the fewest components, none of which are prior US GDP: 2 Canada GDP and 1 UK.

One thing is clear, there is feedback amongst all nations GDP.

Another is that although Canada appears many times in each nations GDP, its coefficients are offsetting positives and negatives.

US does not look at its previous GDPs but to its 2 largest trading partners for prediction.



Above are very crude plots of historical GDP and 8 quarters of forecasts from Jul 2011 until July 2013.